

ORACLE®

SOFTWARE POWERS THE INTERNET

DOEHRS Data Repository

Preliminary Design Review



April 26, 2000

Outline

- Introduction Duane Nathaniel
- Project Overview Duane Nathaniel
- Approach Sharon Hockenberry
- Areas Reviewed John Heaton

Introduction

- Purpose of meeting
 - To offer preliminary findings to the DOEHRs community as a checkpoint.
 - Review of approach, deliverables and methodology.
- Housekeeping
 - Break every hour.
 - Interactive session

Project Overview - Members

DOEHRS Functional Representatives

- LTC Theresa Schulz
- Leeann Domanico
- Brenda Wolbert
- Jeff McClafin
- Tom Helfer

SAIC Personnel

- Gwo Yang
- Maria Cavallaro
- Roger Menzel
- Bob Eek
- Dave Walters
- Anne Mackin

Oracle Project Team

Sharon Hockenberry, Bill Lupton, Isaac Olasupo,
Teresa Mullen, Joe Burke, John Heaton

Project Overview - Scope

- Provide Physical Data Model for DOEHRSD
Data Repository
- Capture Data Elements From HC & IH
Applications
- Provide Designs in Oracle Designer
- Complete By 5 May 2000

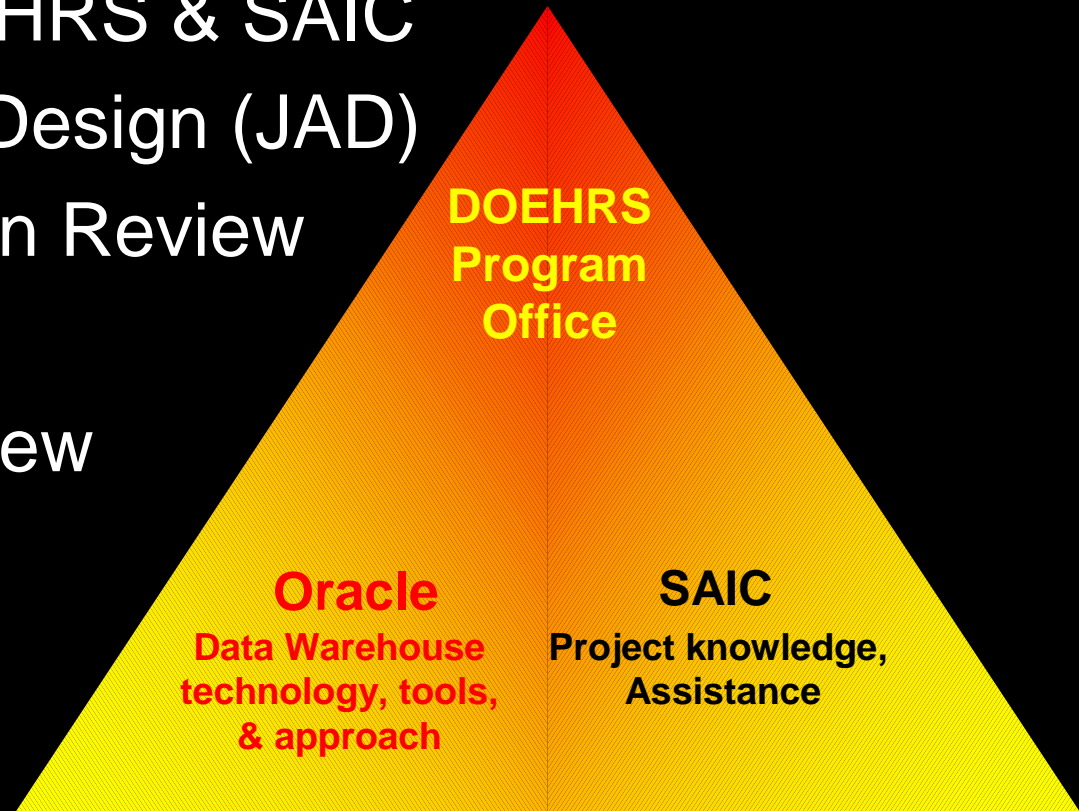
Project Overview - Deliverables

contd

Proposed Deliverable	Mapped Deliverable
Initial Technical Architecture	Technical Architecture
System Interface Requirements	Existing System Interfaces
Business Process Model	Source System High Level Process Flow
Business Function Model	Data Acquisition Strategy
System Data Model	Data Acquisition Requirements
System Process Model	Data Acquisition System Process Model
Logical Database Design	Logical Database Design
Index Design	Index Design
Physical Database Design	Physical Database Design
	Information Requirements
	Engagement Summary & Recommendations
System Function Model	

Approach – Team Integration

- Integrated Team Approach
- Inputs From DOEHRs & SAIC
- Joint Application Design (JAD)
- Preliminary Design Review
- Documentation
- Final Design Review
- Project Sign off



Approach - Initial Assessment

contd

- No “true” data repository component
- Current model designed for reporting
- Current data repository architecture inefficient & not scalable
- ETL design custom coded & manual
- Data transformation limited & sequential
- Few shared business entities between HC & IH models
- Harness existing model supplemented by business requirements to develop the next version.

Approach - Methodology

contd

Deliverable	Description
Information Requirements	The information necessary for the business and its decision support operations. These are high-level needs for the overall enterprise.
Source System High Level Process Flow	Describe the systems from which data is being pulled as well as the type of data, format of data, amount of data and frequency.
Technical Architecture	The architectural design for the enterprise data repository.
Existing System Interfaces	Business Process Flows for data repository.
Data Acquisition Strategy	Approach for the ongoing integrity of the data in the warehouse.
Data Acquisition Requirements	The architectural design for the enterprise data repository.
Data Acquisition System Process Model	Technical movement of data from source systems to target systems.

Approach - Methodology

contd

Deliverable	Description
Logical Database Design	The logical, information model representing a full and detailed definition of the structure of all the data available in the repository.
Physical Database Design	The schema designs, data partitioning, summary and fact tables for the data repository.
Index Design	The specifications for indexing the database.
Engagement Summary & Recommendations	

Approach - Terminology

- **Operational Data Store (ODS)**
 - Houses tactical data from production systems and is subject-oriented and integrated to address operational needs
 - Detailed, current information of a transactional nature, refreshed frequently, and only held for a short period of time (normally)
 - Goal is to provide a tactically-structured, efficient information processing environment to satisfy analysis and reporting capabilities required for the **day-to-day operations** of the business.

- **Data Repository (Data warehouse)**
 - Stores data from operational data sources, using structures organized around major subject areas such as person, MACOM, DOD component, and sample.
 - Goal is to provide an **enterprise** structured, efficient information processing service through increased accessibility, standardization and reliability of timely information.

Approach – System Differences Contd

Data Repository / DW

- Subject Oriented
- Time Variant
- Non-Volatile
- Integrated

OLTP

- Process Oriented
- Real Time
- Volatile
- Usually stand alone

Business Intelligence Systems

- Data Repository
- Data Mart
- ODS

- **Subject Areas**

- Relate to business functions e.g. Audio Tests.
- Very focused for specific function and information.
- Used to manage and report on areas of the business.
- Determine the business key performance indicators to measure performance.

- **Dimensions**

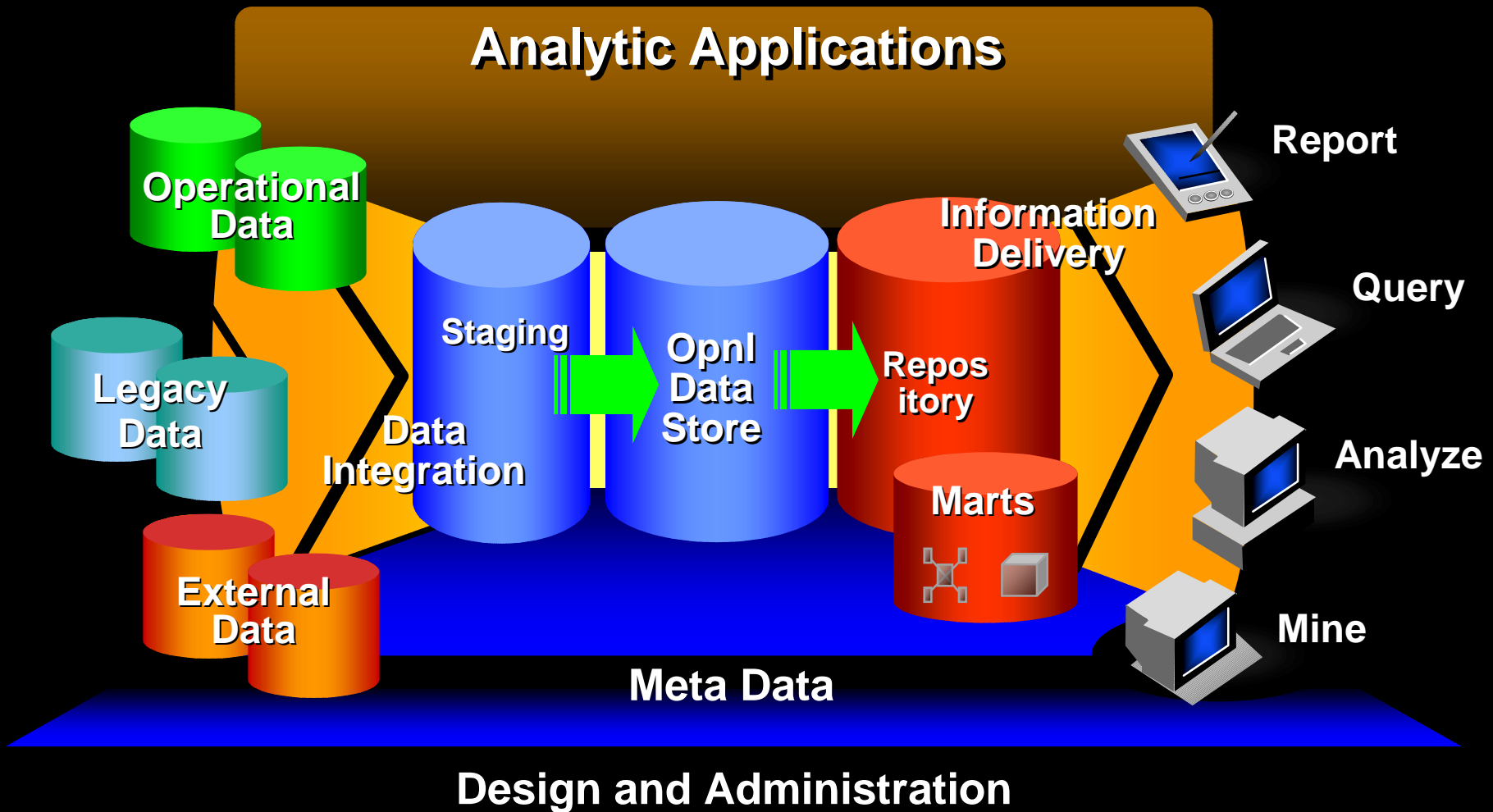
- Describe the business entities of an enterprise, which usually represent hierarchical, categorical information such as time, departments, locations, and products. Dimension tables are sometimes called lookup or reference tables. Dimension tables usually change slowly over time and are not modified on a periodic schedule. They are typically not large, but they affect the performance of long-running decision support queries that consist of joins of fact tables with dimension tables, followed by aggregation to specific levels of the dimension hierarchies

- **Facts**

- **Describe the business transactions of an enterprise. Fact tables are sometimes called detail tables. The vast majority of data in a data warehouse is stored in a few very large fact tables. They are updated periodically with data from one or more operational online transaction processing (OLTP) databases. Fact tables include measures such as exposures and audio test.**

Approach - Oracle Warehouse

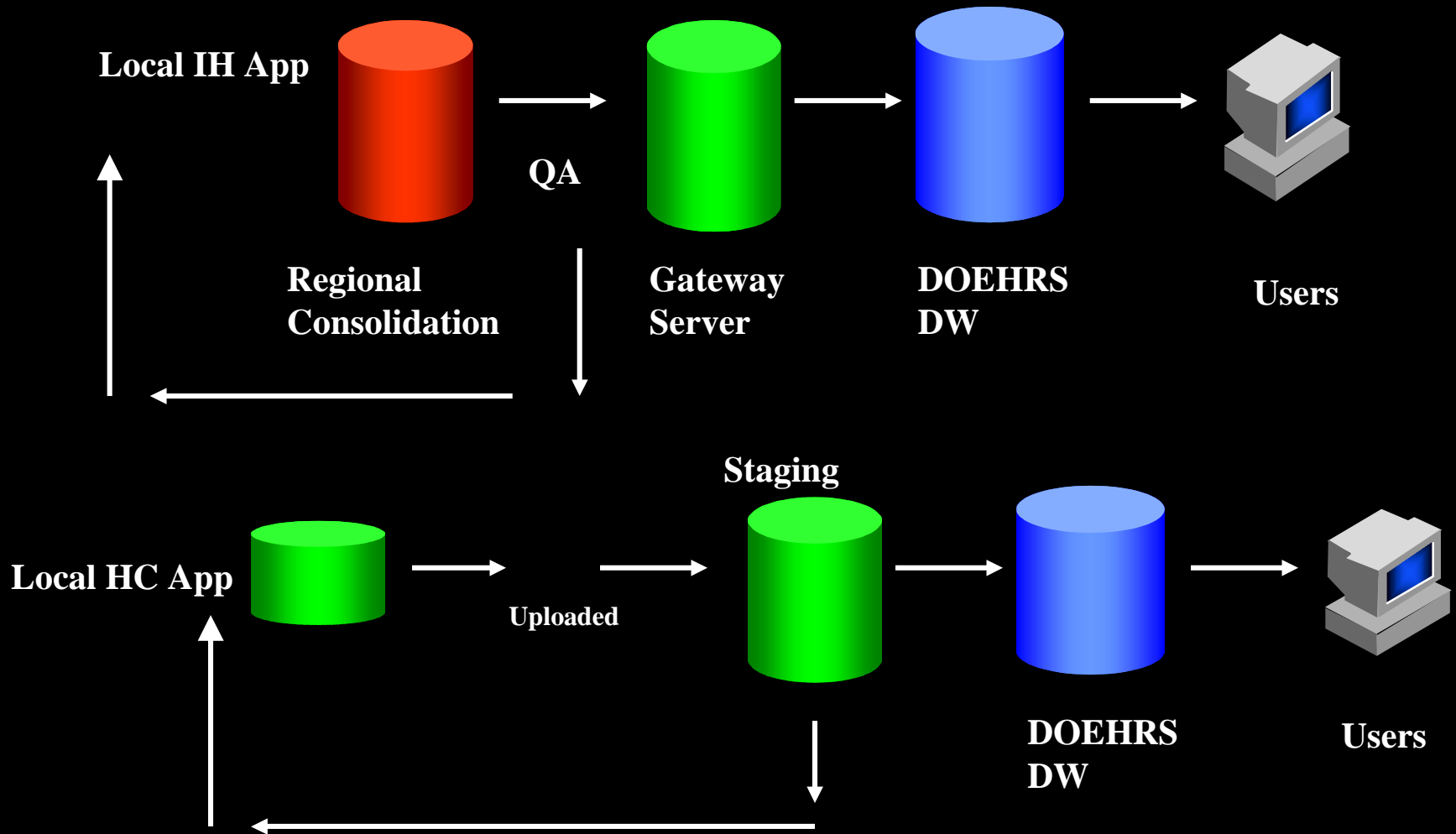
contd



Areas Reviewed

- **Business Process Flow**
- **Technical Process Flow**
- **Technical Architecture**
- **ETL Strategy**
- **Logical / Physical Design**

Areas Reviewed - Business Process Flow



Areas Reviewed - Business Process Flow contd

- Findings

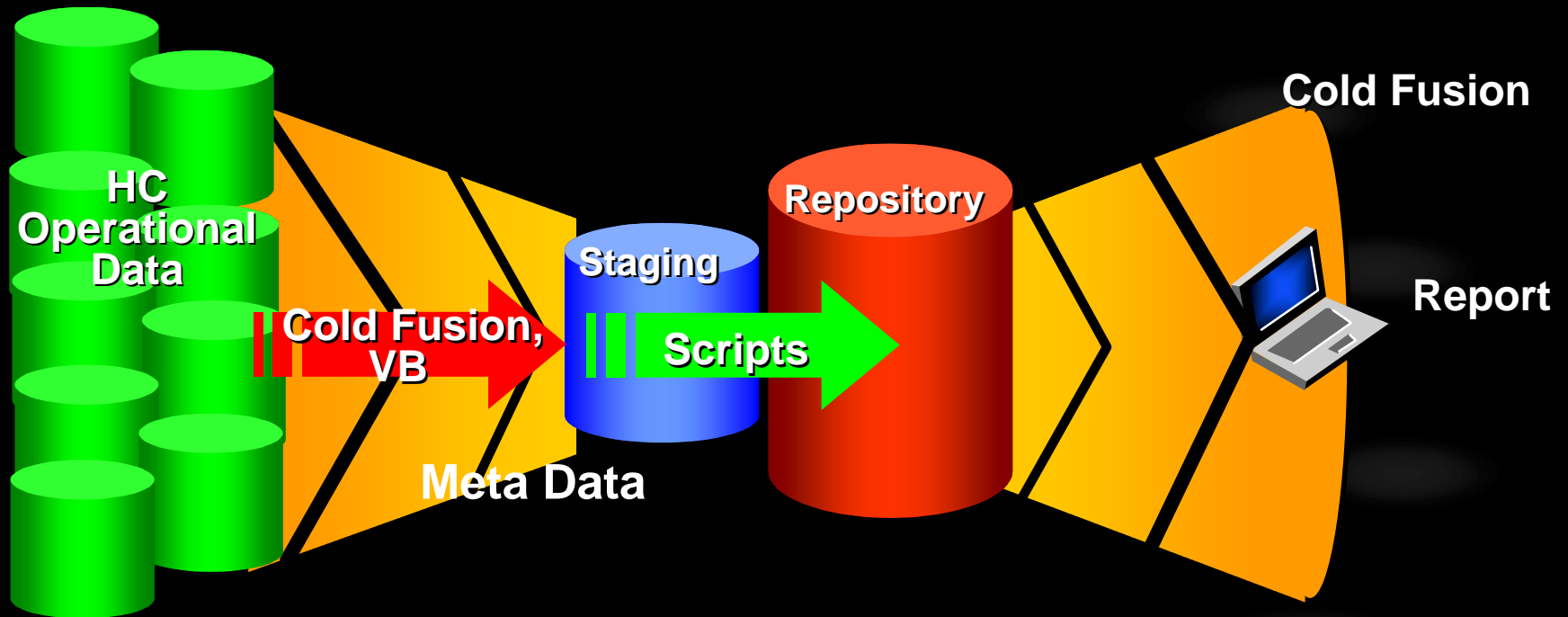
- DR implemented a process to collect HC data.
- Utilizing available resources.
- Consolidation of IH data before passed to DR.
- Manual process flows.

- Recommendations

- Regulate HC process flow.
- Validate IH process flow when production.
- Check for outstanding feeds to DR.
- Automate process flows.

Areas Reviewed - Technical Process Flow

contd



Areas Reviewed - Technical Process Flow **contd**

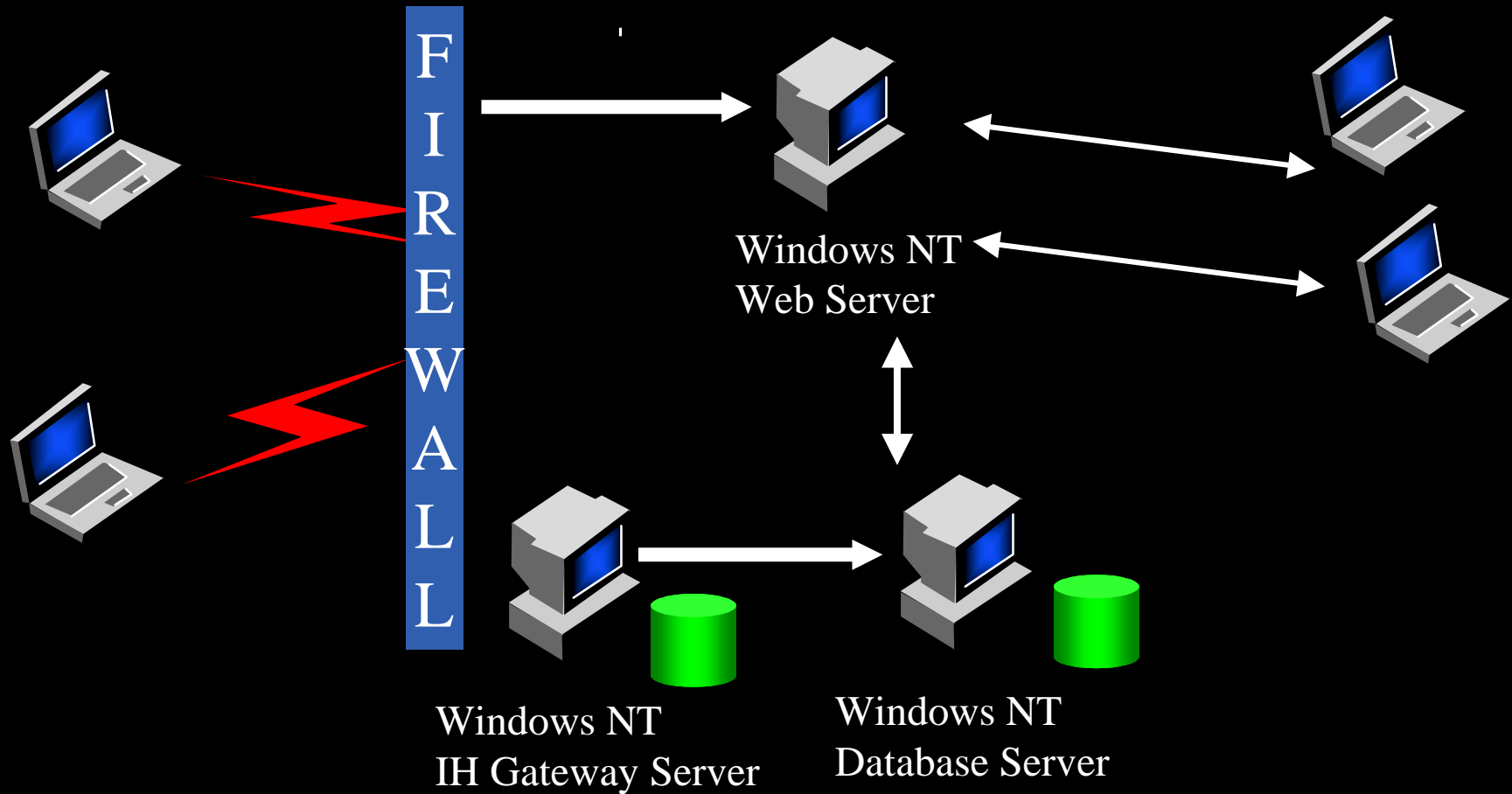
- Findings

- Custom Scripts.
- Push strategy of information for HC.
- Push / Pull strategy of information for IH.
- Minimal use of staging areas.
- Minimal use of meta data.
- Serial and row processing.
- Architecture not scaleable.
- No consolidation of HC information before DR.
- No automated uploads.

- Recommendations

- Change Push to pull strategies.
- Use bulk and parallel loading techniques.
- Automate data loading.
- Enhance technical and business meta data for load, warehouse and query management.
- Utilize additional features of RDBMS.
- Consolidate HC information before DR.
- IH technical process flow vague.
- Replication Management for HC and IH.

Areas Reviewed - Technical Architecture



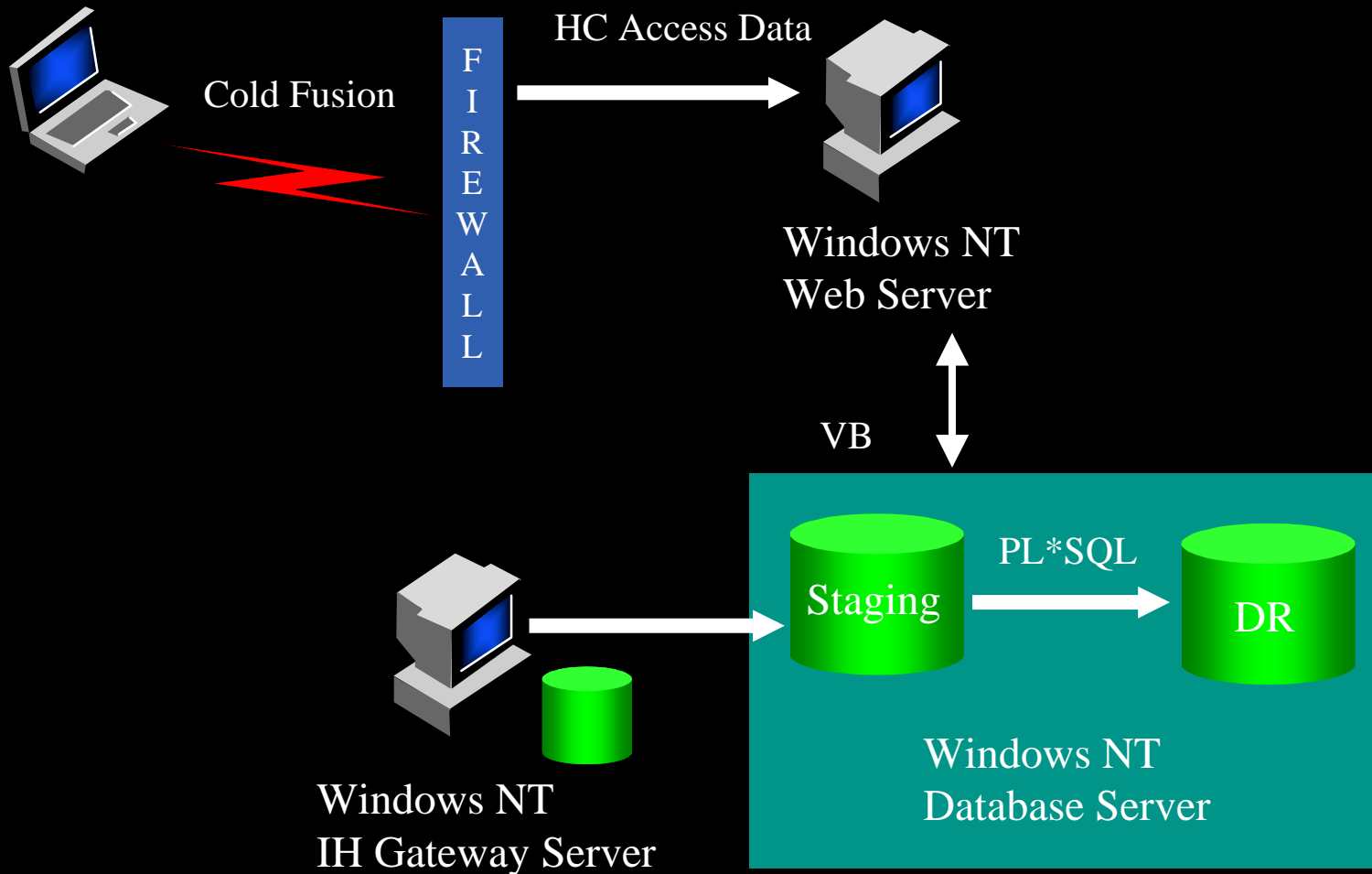
- Findings

- User Community approx. 2500.
- Web Server and Database Server separate (Three Tier architecture).
- Cold Fusion reporting and load management.
- Ability to collect ACCESS data from distributed sites.
- Scalability / performance risk with Windows NT / 2000 for projected user base.
- Performance metrics vague.
- High end NT hardware.
- Backup strategy in place.

- Recommendations

- Re look at technical architecture for scalability and performance (RAID setup, database configuration, potential use of other operating systems, clustering)
- Use of Oracle RDBMS options to provide scalability and performance (Partitioning and Parallel Server).
- Inclusion of additional tools to manage environment.
- Add batch capabilities for reporting.

Areas Reviewed - ETL Strategy



Areas Reviewed - ETL Strategy

contd

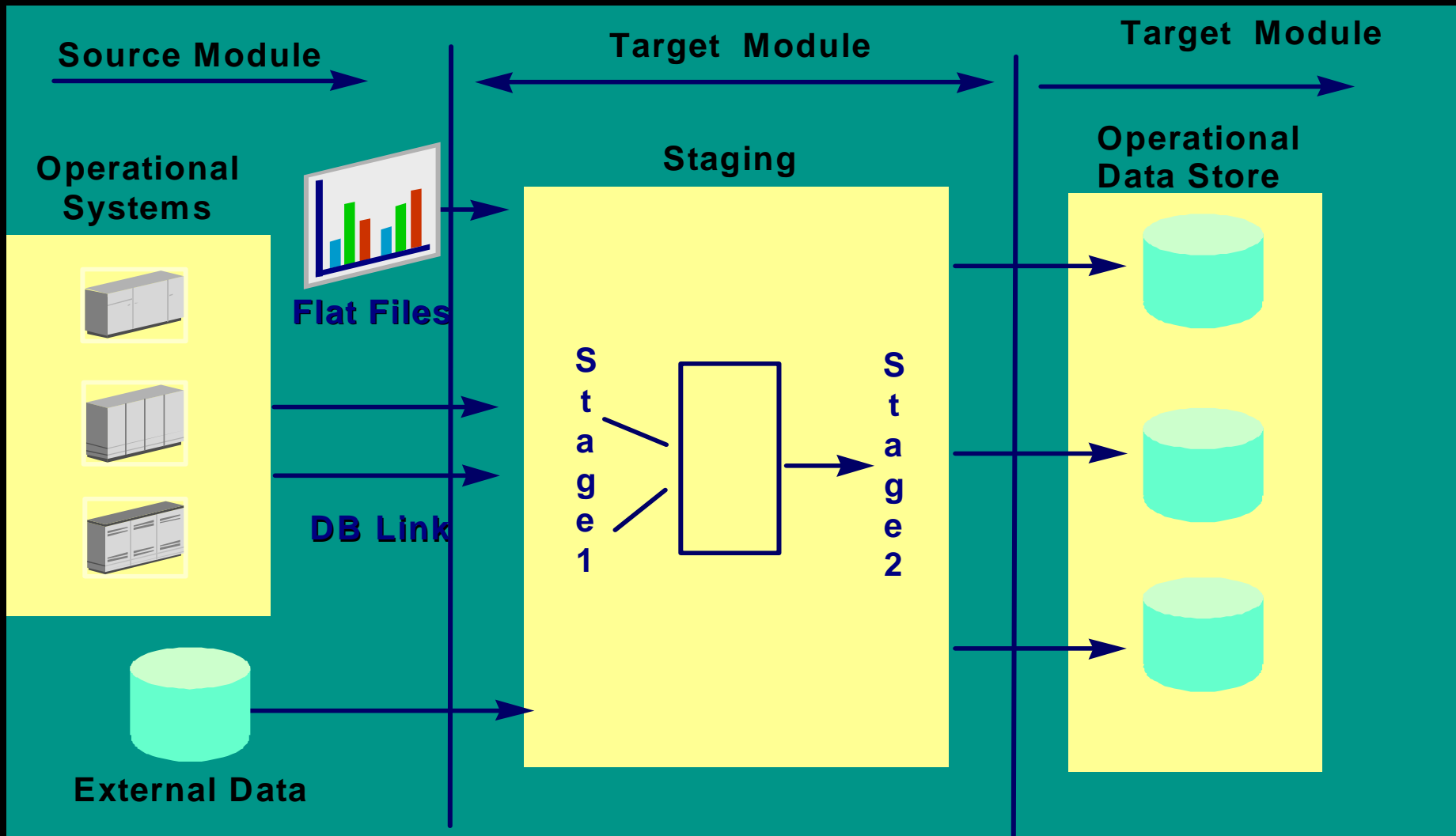
- Findings

- Custom Code.
- Sequential.
- Row level operations.
- Manual load process.
- Different processes for historical and current data.
- Individual summaries for each report requirement.
- PL*SQL is scalable.
- VB is not scaleable.
- Code management is in place.

- Recommendations

- Use Parallel bulk loads.
- Include versioning.
- Include person consolidation.
- Undergo Source – Target mappings to identify business rules to ensure data integrity, quality and completeness.
- Define error processing strategy.
- Create at least a dual staging area.
- Change HC app to export source information.
- Automation of loads
- Common load processes.
- Utilizing tools that can reuse existing investment.
- Redesign to utilize RDBMS features and to increase scalability, performance and audit capabilities.
- Archival strategy.

Areas Reviewed – Proposed ETL Strategy



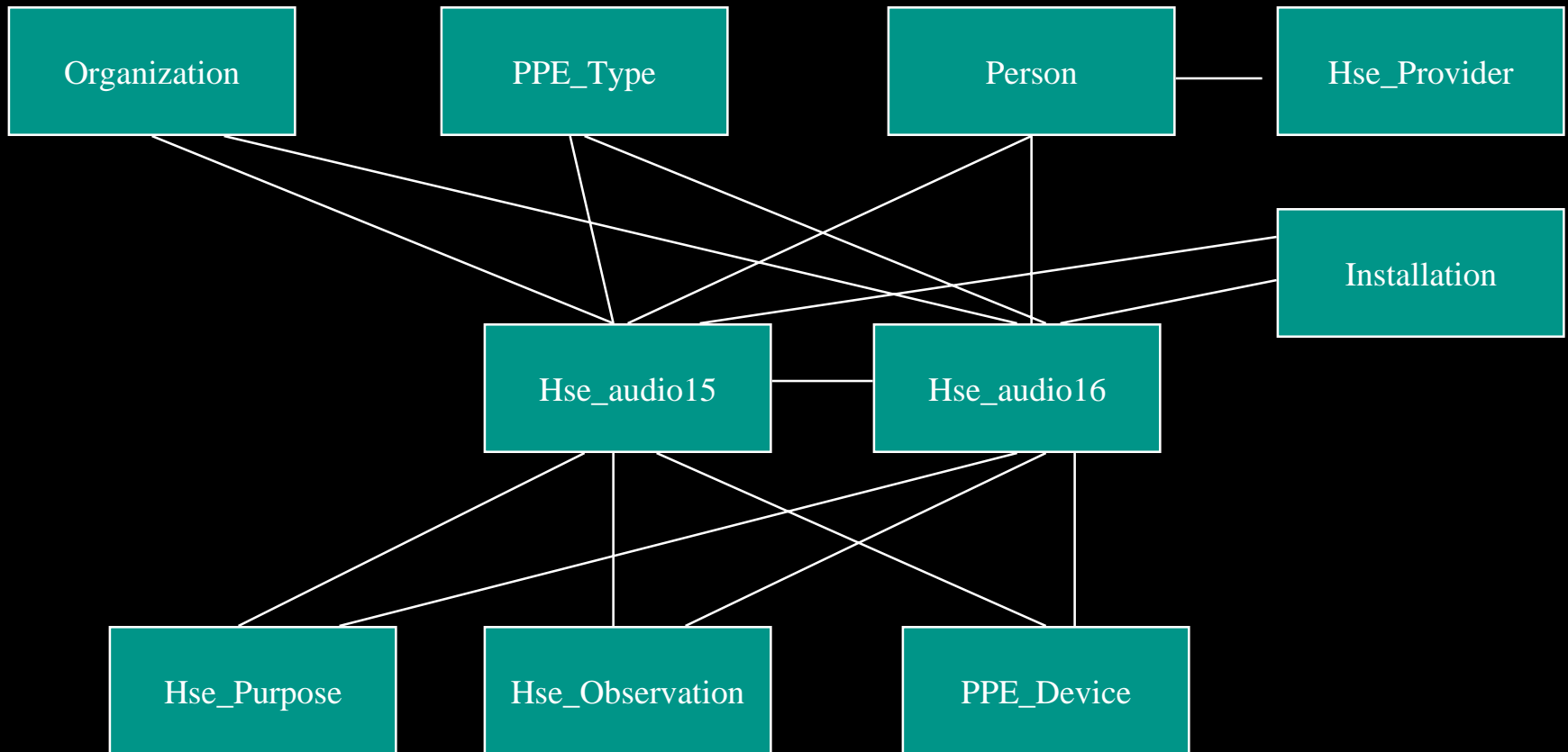
Areas Reviewed – Logical Design Process

- Hybrid approach – Top down and bottom up.
- Joint Application Design sessions for review of model and collection of requirements.
- Consultation and use of existing models.
- Report requirements.
- Design walkthroughs to define rules for DR.

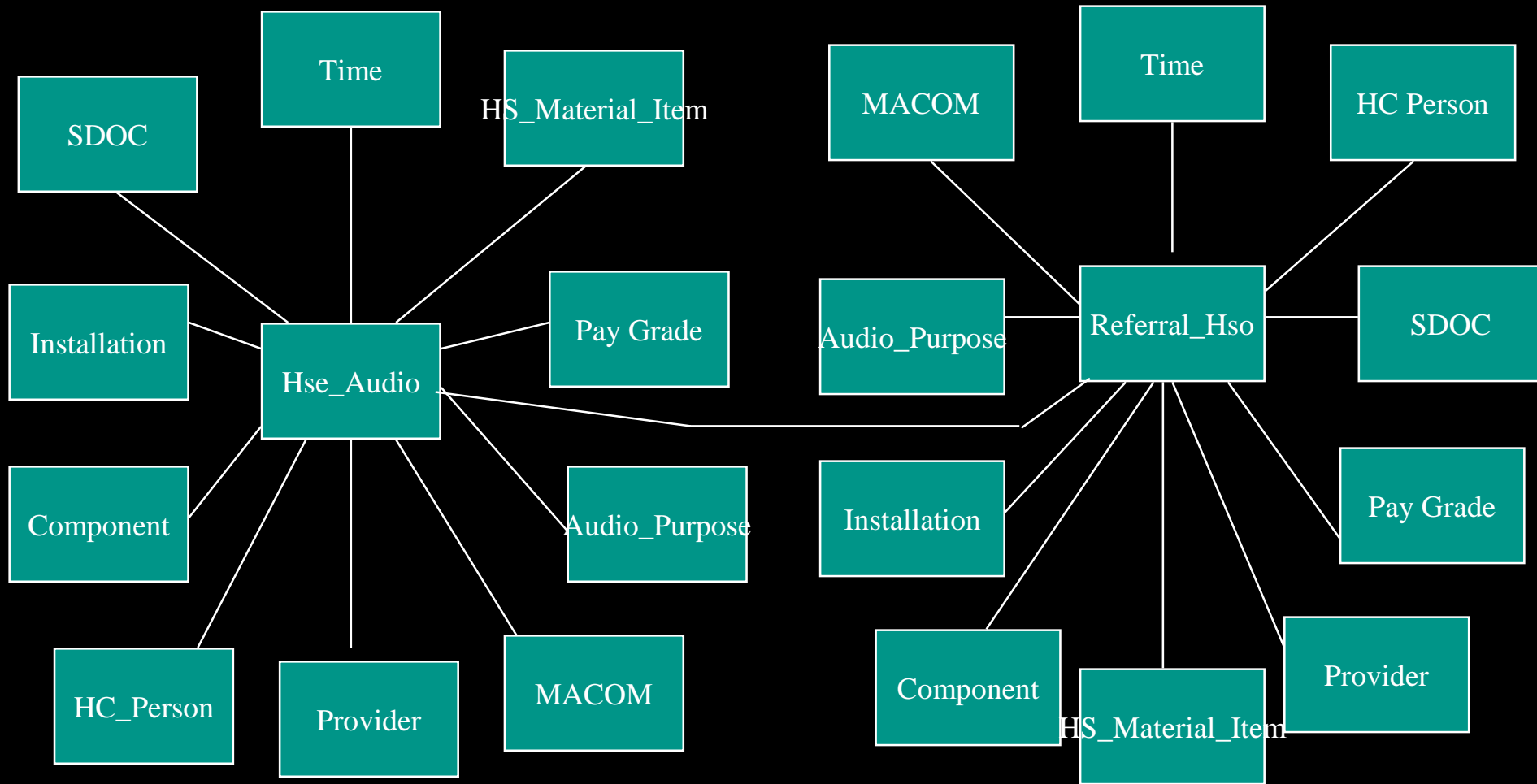
Areas Reviewed – Logical Design Subject Areas

- Current Iteration
 - Audio Tests – HC
 - Referrals – HC
 - Exposures – IH
 - Surveys – IH
 - Samples – IH
- Next Iteration
 - Base camp assessments - IH
 - Workload Requirements - IH
 - Indoor air quality - IH
 - IH Equipment - IH
 - Ergonomics - IH

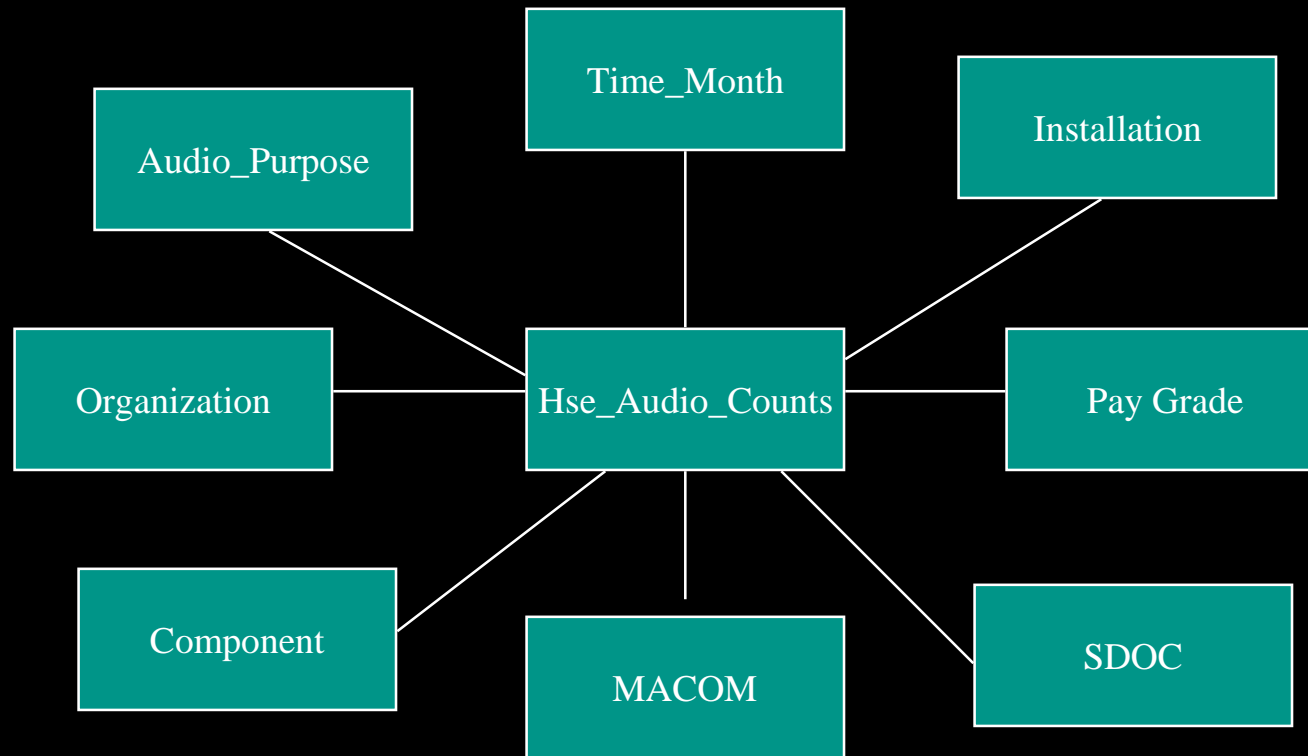
Areas Reviewed – Current HC Logical Design ODS



Areas Reviewed – Proposed HC Logical Design ODS



Areas Reviewed – Proposed HC Logical Design DR



Areas Reviewed – HC Deltas

- Consolidated Audio15, Audio16, PPE_TYPE, Hse_observation
- Broke out SDOC, DoD / Service Component, MACOM, Pay Grade.
- Added Time dimension.
- Renamed dimensions.
- Enhanced hierarchies.
- Defined HC DR model.

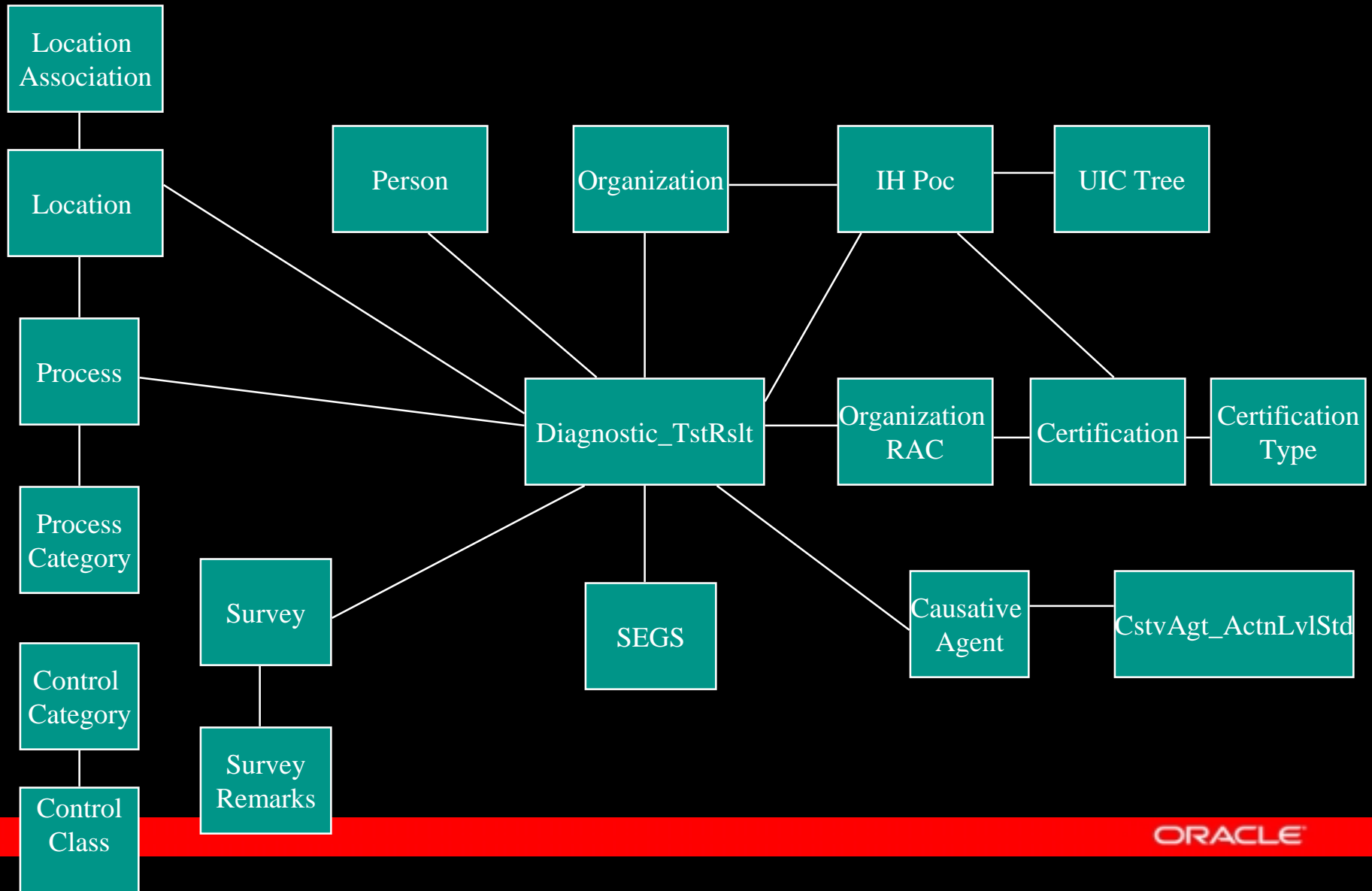
Areas Reviewed – HC Subject Areas ODS

- Subject Areas and Hierarchies
 - Audio Tests
 - Person / Organization
 - SDOC
 - Pay Grade
 - Provider
 - Material Item
 - Service Component / DoD Component
 - Installation
 - Audio Purpose
 - MACOM
 - Time
 - Referrals
 - Same as above

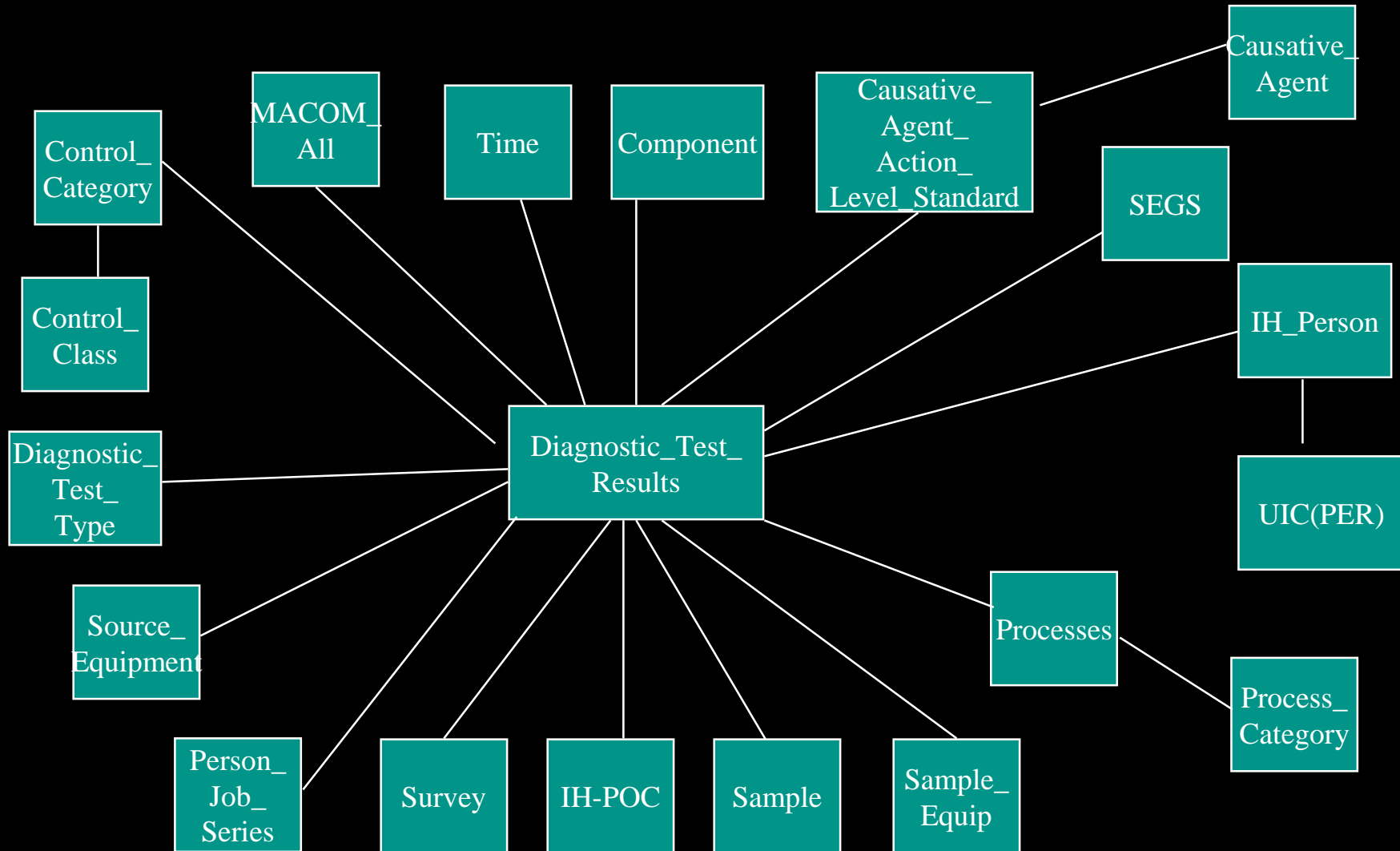
Areas Reviewed – HC Subject Areas DR

- Subject Areas and Hierarchies
 - Audio Counts
 - Organization
 - SDOC
 - Pay Grade
 - Service Component / DoD Component
 - Installation
 - Audio Purpose
 - MACOM
 - Month
 - Three additional summary tables have been defined

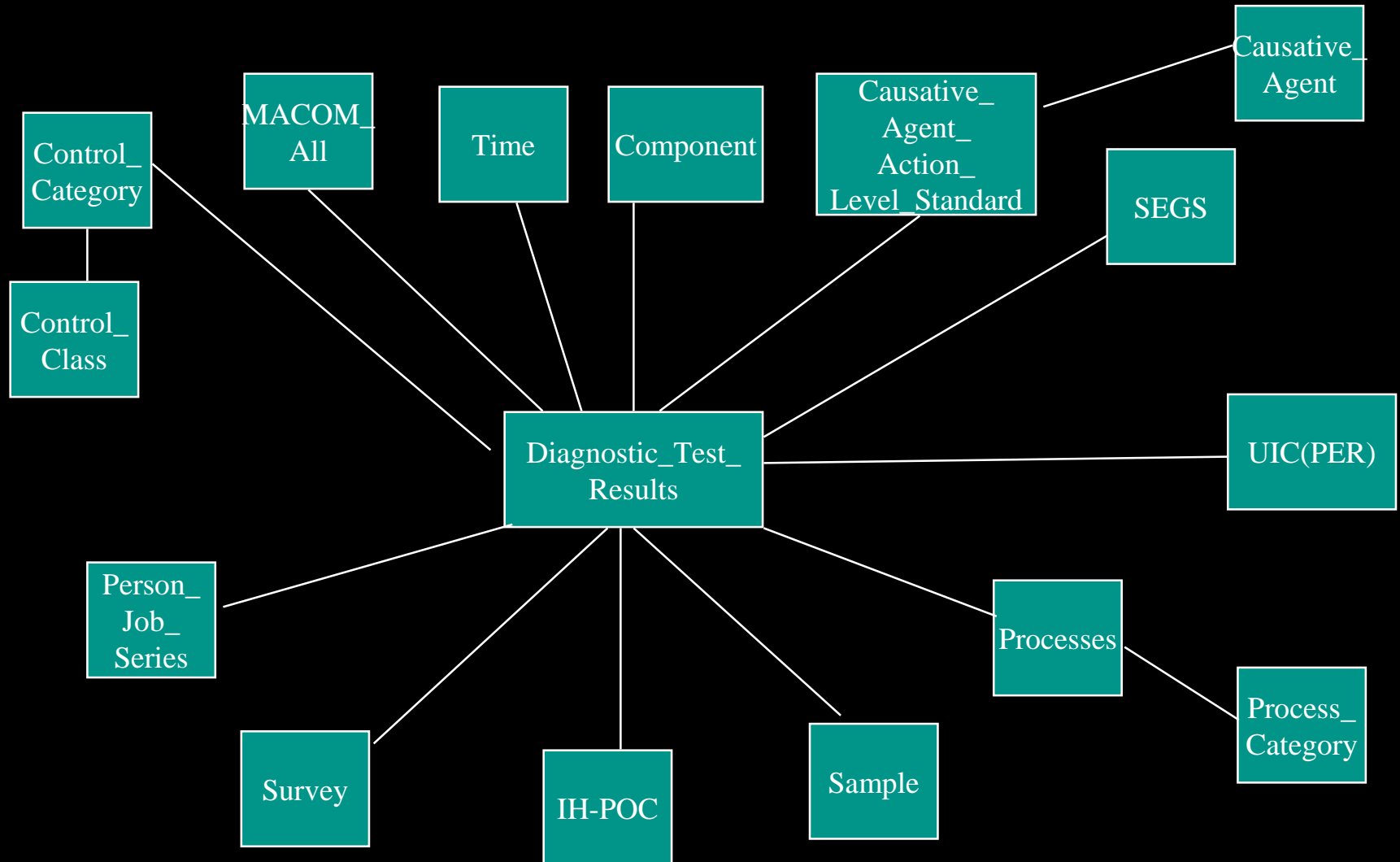
Areas Reviewed – Current IH Logical Design ODS



Areas Reviewed – Proposed IH Logical Design ODS



Areas Reviewed – Proposed IH Logical Design DR



Areas Reviewed – IH Deltas

- Consolidated Analysis Methods, Diagnostic Test Results
- Broke out Job Series, DoD / Service Component, MACOM.
- Added Time and Sample dimensions.
- Enhanced hierarchies.
- Defined IH DR model.

Areas Reviewed – IH Subject Areas ODS

- Subject Areas and Hierarchies
 - Diagnostic Test Results
 - MACOM / All MACOMS
 - Service Component / DoD Component
 - Time
 - Causative Agent Action Level / Causative Agent
 - SEGS
 - Person / UIC
 - Process / Process Category
 - Sample Equipment
 - Sample
 - IH_POC
 - Survey
 - Job Series
 - Source Equipment
 - Diagnostic Test Type
 - Control Category / Control Class

- Subject Areas and Hierarchies
 - Certification
 - IH_POC
 - Service Component / DoD Component
 - Certification Type
 - Person
 - Time
 - Survey Remarks
 - Survey
 - Causative Action Level Std / Causative Agent
 - Process / Process Category
 - Control Category / Control Class
 - Time

- Subject Areas and Hierarchies
 - Selected UIC
 - Service Component / DoD Component
 - IH_POC
 - Time
 - UIC

Areas Reviewed – IH Subject Areas DR

- Subject Areas and Hierarchies
 - Diagnostic Test Counts
 - MACOM / All MACOMS
 - Service Component / DoD Component
 - Time
 - Causative Agent Action Level / Causative Agent
 - SEGS
 - UIC
 - Process / Process Category
 - Sample
 - IH_POC
 - Survey
 - Job Series
 - Control Category / Control Class

Areas Reviewed – IH Subject Areas DR

- Subject Areas and Hierarchies
 - Organization Risk Assessment Code
 - SEGS
 - UIC
 - Service Component / DoD Component
 - MACOM
 - Time
 - IH_POC
 - Causative Agent
 - Process / Process Category

Areas Reviewed – Reasons for Changes IH and HC

- Added flexibility for reporting.
- Performance enhancements.
- Sharing of dimensions across IH and HC to enable cross functional reporting.
- Common information grouped via subject areas.
- Allows for incremental build of DR.

Areas Reviewed – Commonalities IH & HC

- Person
- MACOM
- Service component / DoD component.
- Time
- Audio Tests

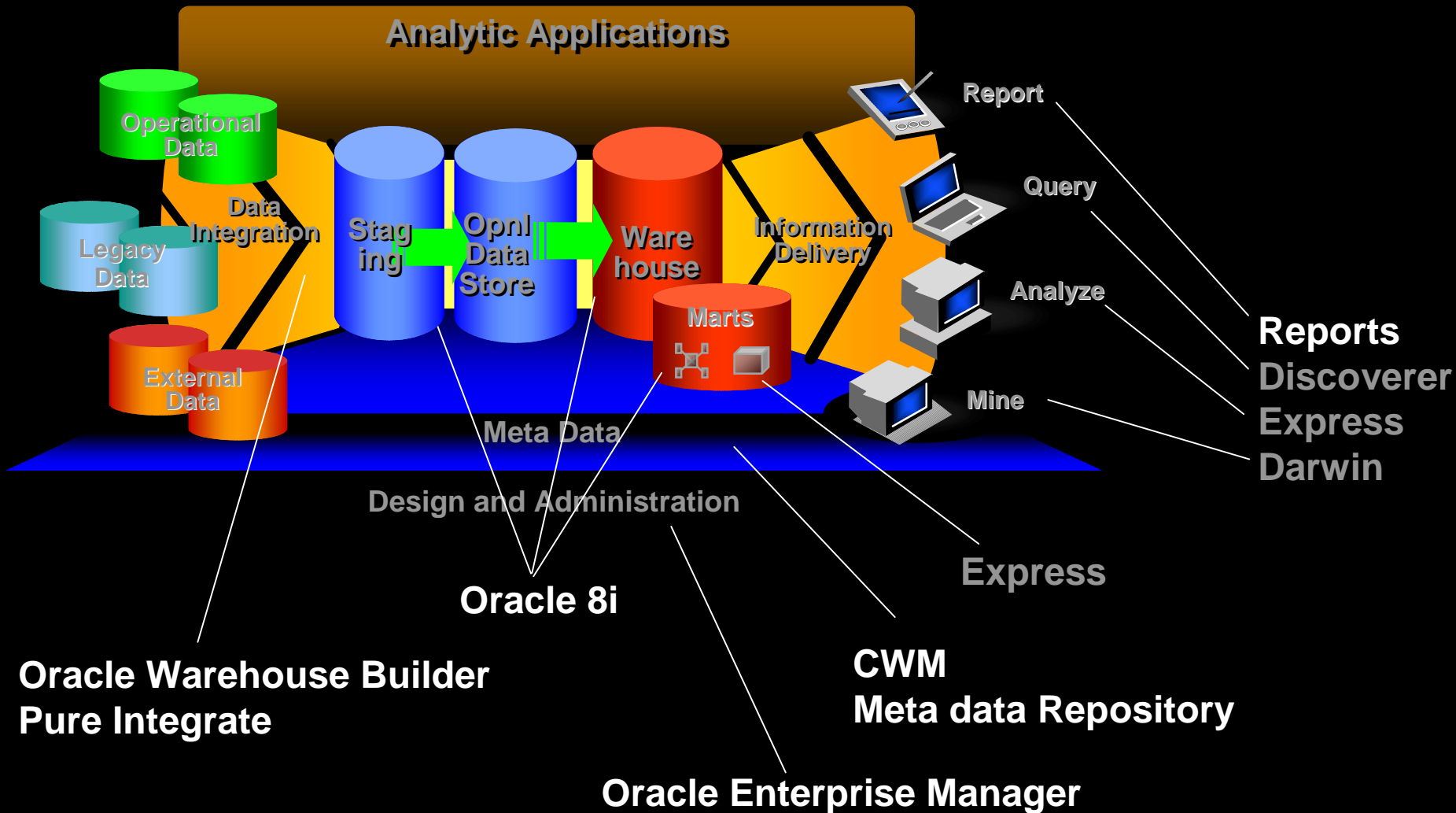
Areas Reviewed – Physical Design

- Denormalize hierarchies.
- Integrate snow flaked entities.
- Partition based on time and MACOM.
- Add any additional columns for auditing purposes.
- Design indexes.

Data Warehouse Architecture Prerequisites

- Requirements
 - Consolidated Person information
 - ETL tools which can reuse current PL*SQL functions
 - Automated Jobs scheduling
 - Load Management
 - Summary Management
 - Query Management
 - Warehouse Management
 - Meta Data Management
 - Batch Reporting
 - ADHOC Reporting
 - Analytical Analysis

Approach Tools



Next Steps

- ETL Mapping and documentation of business rules for data quality, consistency, integrity and completeness.
- Tools selection to minimize custom code, maximize automation and reuse.
- Data warehouse architecture redesign for performance, scalability and maintainability.
- Design and build of integrated ETL strategy to support IH, HC and potential future sources.
- Gather requirements for the logical design of incremental phases.

ORACLE®

SOFTWARE POWERS THE INTERNET